FEBS Letters 582 (2008) 1220–1224

Interaction networks for systems biology

Samuel Bader, Sebastian Kühner, Anne-Claude Gavin*

EMBL, Structural and Computational Biology Unit, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

Received 28 January 2008; accepted 8 February 2008

Available online 20 February 2008

Edited by Robert Russell and Patrick Aloy

Abstract **Cellular functions are almost always the result of the coordinated action of several proteins, interacting in protein complexes, pathways or networks. Progress made in devising suitable tools for analysis of protein–protein interactions, have recently made it possible to chart interaction networks on a large-scale. The aim of this review is to provide a short overview of the most promising contributions of interaction networks to human biology, structural biology and human genetics.**

Keywords: Interactome; AP–MS; Yeast two-hybrid; Protein complex; Systems biology; Protein–protein interaction

1. Introduction

Biology relies on the concerted action of a number of biomolecules organized in pathways or networks. Molecular biologists have traditionally studied the interactions and relative influence of biomolecules in focused, one at a time studies. These approaches have contributed insight on limited numbers of signaling pathways and on the cellular or physiological functions of proteins.

More recently, advances in high throughput methods make it possible to study the behavior and attributes of biomolecules that make up entire systems. This recent interest in whole systems comes from the belief that systems have functions that none of the entities of the systems have, and that "the total is more that the sum of its parts". The rules that govern the behavior of biological systems are currently the focus of intense research in the field of Systems Biology. The resulting models are expected to be predictive of different healthy and pathological conditions. They might provide synthetic biologists with the general principles for the (re)engineering of biological systems for particular purposes [1,2].

In this blooming field, systems-wide analyses of protein–protein interactions have taken center stage. A number of strategies have been applied to the charting of protein–protein interactions on a large-scale. They include, the yeast two-hybrid system [3,4] Affinity Purifications/Mass Spectrometry (AP–MS) [5–9] and in silico prediction (reviewed in [10]). The basic principles behind these approaches as well as their advantages and limitations have been the subject of extensive review (for example [10,11]) and will not be discussed here. Rather, we will report on the recent applications of these methods to genome-wide screens.

The field is rapidly maturing and the outcomes and progresses made are clearly visible. Even if the coverage remains quite low and the resulting networks are usually characterized by a high rate of false negatives (only about 15% of all possible interactions have been charted so far [12]), the first cartographies of several pathways and networks that map behind main human pathologies have already emerged. In model systems, global, genome-scale, protein–protein interaction screens have provided a molecular framework for the interpretation of very simple genetic data, such as gene essentiality [13]. A number of databases have been developed that integrate standardized biomolecular interaction data from various origins [14].

2. Modality of protein–protein recognition; structures

Proteins inside the cell do not interact randomly. Protein associations need precise regulation. The spatial and temporal regulation of enzyme activities through extensive interactions bears remarkable functional relevance. In human, mutations or environmental factors that interfere with protein–protein interaction lead to pathology. This is the case in the Immunodeficiency, Centromeric instability, Facial anomalies (ICF) syndrome, caused by defects in DNMT3B, a DNA methyltransferase. The missense mutations have been mapped not only within the catalytic site but also affect an N-terminal PWWP domain of DNMT3B, involved in protein–protein interactions [15]. Even discrete changes in the affinities between two interacting protein pairs can have devastating consequences. For instance, mutations in the Fibroblast Growth Factor Receptor 2 (FGFR2) that selectively increase the affinity for FGF2 [16] are responsible for the Apert syndrome, characterized by skull malformation, syndactyly and mental deficiency.

A wide variety of modular and specialized binding domains have been mapped that mediate protein–protein recognition [17]. Even though the distinction is not always very strict, interaction domains are thought to generally operate through two mechanisms (Fig. 1). Domains sometimes bind other domains. Typically, this type of interactions involves large binding interfaces provided by the rigid globular domains. This is illustrated by the interaction taking place between Ras and its GTPase activating protein Ras-GAP, which play a crucial role in regulating cellular signal transduction processes [18]. Domain–domain interactions are thus generally characterized by relatively high stabilities and affinities in the low nM to pM range [19,20].

*Corresponding author.
E-mail address: gavin@embl.de (A.-C. Gavin).

0014-5793/$34.00
doi:10.1016/j.febslet.2008.02.015

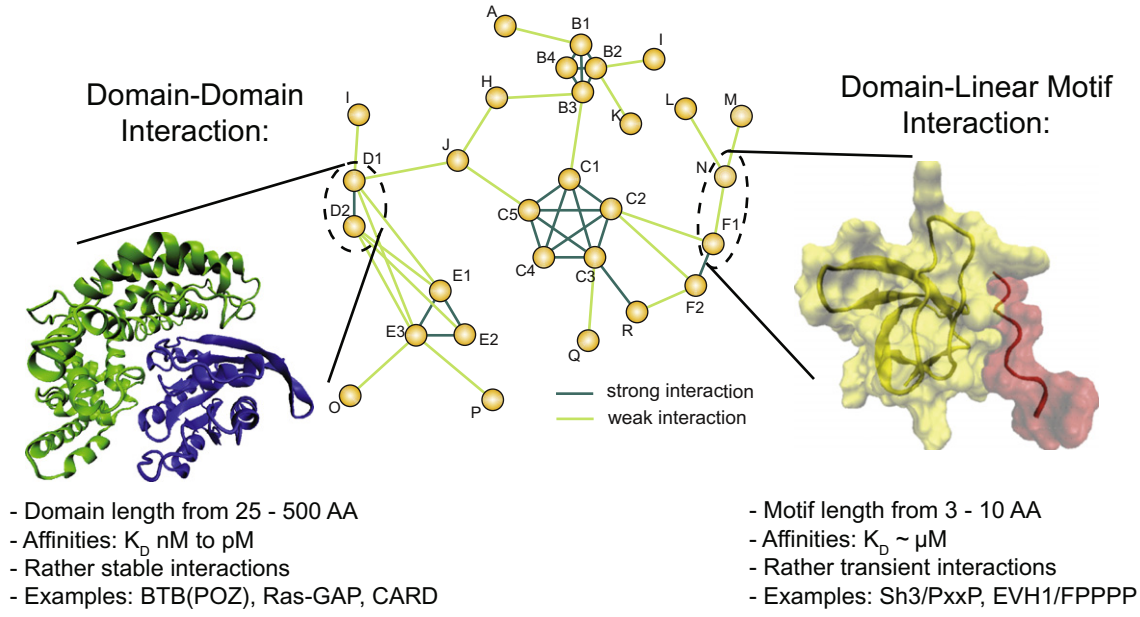

Fig. 1. Protein interactions usually happen through two rather distinct mechanisms. Protein recognition mediated by domain–domain interaction is illustrated on the left part of the figure with the binding of Ras to Ras-GAP (PDB:1wq1). Domain-linear motif interactions are exemplified on the right with the binding of the proline rich ligand RALPPLPRY to a SH3-domain (PDB: 1RLQ). A few characteristic features of each mode of binding are also highlighted. Structures are visualized using the program VMD [58].

In other cases, interactions are mediated by short sequences, typically between 3 and 10 amino acids, present in disordered parts of proteins (Fig. 1). For example, Src Homology 2 (SH2) domains specifically interact with small peptides containing a phosphotyrosyl residue. PDZ domains can also target small (4 amino acid long) consensus binding motifs located at the C-terminus of the interaction partners. These short linear motifs are critical to many biological processes. They often show low affinities (0.5–10 µM) [21,22]. They tend to be mediators in transient interactions, such as the ones seen in cell signaling [23]. Residues of the linear motif can directly contribute the totality of the binding energy. But in many cases, additional indirect induced fit mechanisms are important to stabilize the interaction [24]. The number of known linear motifs capable of engaging in an interaction is still relatively small (a few hundreds). They have traditionally escaped the detection by classical sequence alignment algorithms. The recent availability of several large interaction networks have open the way to more systematic approaches for finding protein linear motifs that mediate protein–protein interactions [22,25]. These motifs have recently attracted much attention as targets of small molecules that disrupt protein–protein interactions (reviewed in [26,27]).

In the majority of the cases the sequence or the structural determinants responsible for the specificity and the precision of the recognition remain unknown. Methods such as mutation (alanine) scanning, or the monitoring of binding affinities to peptide libraries have provided and are still providing, very interesting and valuable information regarding the sequences involved. Structural determinants are certainly more difficult to capture. Recent strategies have been developed that aim at the systematic structural resolution of macromolecular assemblies. These strategies integrates large datasets on the biochemical composition of protein complexes with data on their low-resolution structure by Electron Microscopy (EM) and the X-ray crystallography of individual interacting domains or proteins [28–31]. Ultimately, these efforts may contribute to the resolution of the protein assembly code that converts the genome's information into a biologically functional third dimension [32,33].

3. Interaction networks in diseases

Protein interaction networks reveal the connectivity of a proteome in a given cellular context. They reflect a particular cellular status. Protein interaction networks are dynamics and change in time and space to adapt or switch to different physiological conditions. An attractive application is the capturing of the changes in protein connectivity that associate with progression towards diseases (Fig. 2).

A field concerns the study of pathogenic organisms and their interface with host cells. A growing number of interaction networks from many different pathogens have been recently produced. They include several viruses such as the Kaposi Sarcoma-associated Herpes Virus (KSHV), Varicella-Zoster Virus (VZV) and Epstein-Barr Virus (EBV) [34,35]. Interestingly, these viral networks appear to possess distinct overall topologies from the ones of the host cells. Most biological networks follow the so called small-world and scale-free behavior, where only few nodes act as "highly connected hubs" and the majority of the nodes are of low degree (are engaged in only few interactions). This attribute is believed to confer resistance to random attacks but makes scale-free networks extremely susceptible to targeted perturbations. The viral interaction networks do not possess these properties. The current hypothesis is that viral networks, in the absence of a host, are incomplete. Consistent with this view is the observation that upon docking on the cellular network, the viral networks assimilates the properties of the host one [34,35]. As they become more com-

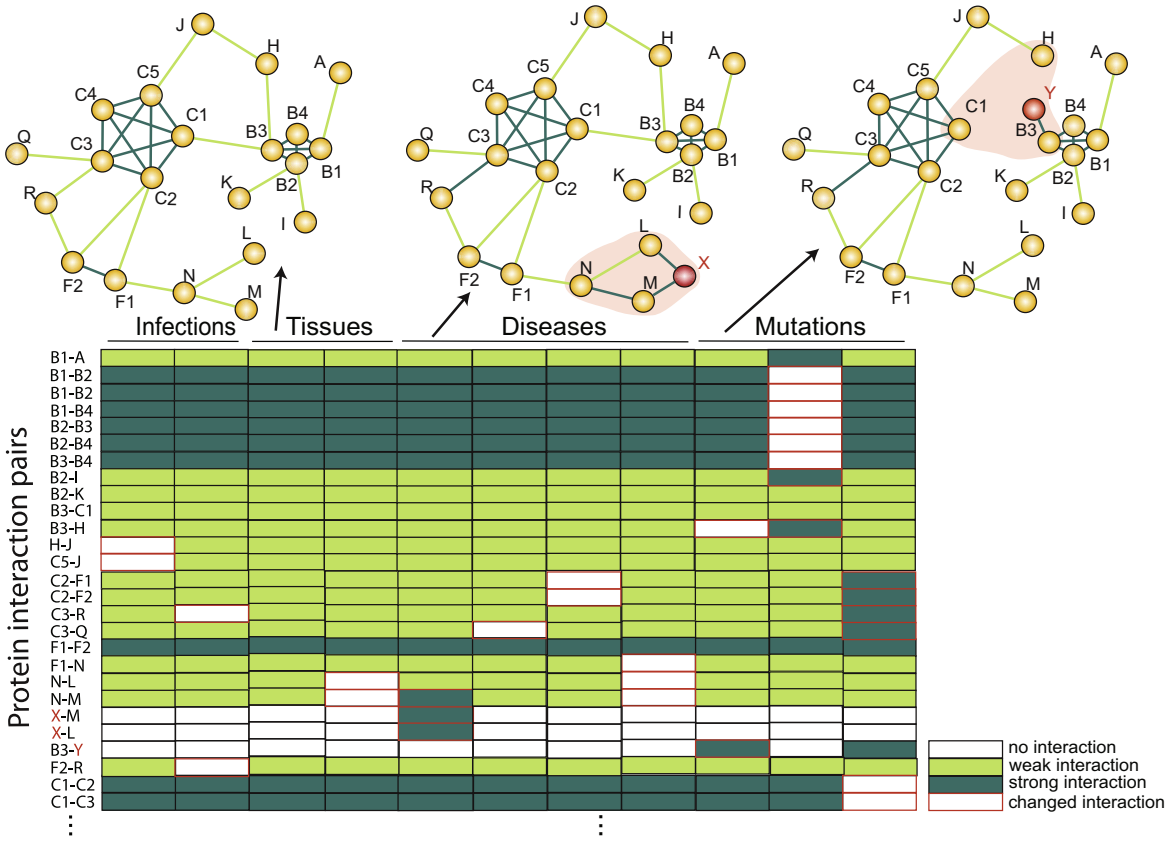

Fig. 2. Protein interaction networks from different cellular status. Different networks illustrative of different cellular conditions are represented as graph on the top part of the figure. The light green lines represents weak interaction, dark green lines are strong ones. Red circles illustrate proteins that are not present in all networks and red halos highlight the changed part of the interactome. The heat map depicts the interactions happening in each cellular state. Boxes are colour-coded according to the strength of the interaction.

plete host-pathogen interaction networks are expected to bring a broader view on how viruses use the cellular machinery to their own purpose. New opportunities for small molecule inhibitors may emerge.

More challenging approaches concerns the charting of the dynamics changes happening during progression towards various diseases status. The analyses have been traditionally limited by the intrinsic difficulties in charting the dynamics properties of protein–protein interactions on a large-scale (reviewed in [36]). Many of the currently used approaches for the analysis of protein–protein interactions imply the expression of proteins under non-physiological conditions (for example in ex vitro systems, such as yeast two-hybrids) or in non-synchronized, heterogeneous populations of cells (for example for biochemical approaches such as AP–MS). The spatial and temporal regulations are usually lost.

Early attempts to fill this gap come from studies in a model organism, *Saccharomyces cerevisiae*. The dynamics of interactions was captured by the superimposition of the temporal changes in gene expression that associate with different cellular conditions on static protein interaction networks [37–39]. Recently, these approaches have been applied in human to the modeling of aging [40] or drug addiction [41]. An aging interaction network was built by the integration of the transcriptional changes observed in tissues derived from young and old humans with protein–protein interaction networks. Interestingly, genes that are differentially expressed during aging seem to often represent key regulatory nodes, probably important for overall network stability [40].

4. Interaction networks as molecular frame for genetic data

Proteins often serve more than one cellular function. One manifestation of this so called pleiotropy is the observation that different mutations in a single gene cause multiple, significantly different phenotypes [42]. Molecular mechanisms, such as alternative splicing and post-translational modifications, probably account for pleiotropy. Gene pleiotropy might also rely on the tendency of proteins to associate with different partners in different cellular contexts and, as a result, exert diverse functions [43]. In human, good illustrations are mutations in the Xeroderma Pigmentosum group D-complementing protein (XPD) gene that cause diseases with different symptoms: Xeroderma Pigmentosum (XP) or Trichothiodystrophy (TTD). XPD is a subunit of the TFIIH basal transcription factor complex implicated in both transcriptional regulation and DNA repair through the selective recruitment of specific factors [44]. The current hypothesis is that different mutations in XPD affect either the DNA repair function of TFIIH (XP results) or its transcriptional role (TTD results) [45]. More generally, the degree of connectivity of a protein positively correlates with the number of traits it influences or its degree of pleiotropy [46].

Genome-wide analysis of protein complexes in yeast shows that almost 40% of the proteins are part to more than one protein complex [47]. This raised the view that protein multifunctionality may be a more general attribute than anticipated. In higher eukaryotes, interaction networks may not only provide a molecular frame for the explanation of genetic traits, like pleiotropy, but are also expected to contribute to the selection of more specific and "safer" drug targets.

It is also relatively common that mutations in different genes sometimes lead to similar or related phenotypes. Generally it is assumed that this reflects disruptions in proteins that participate in a common interaction network, such as the different subunits of a multiprotein complex or proteins that function at different steps of a signaling or biochemical pathway [48]. In human, a striking example is provided by the nine genes associated with the Fanconi Anaemia (FA) syndrome that cooperate in a common network, the FA/BRCA2 (Breast Cancer type 2) pathway involved in DNA repair. Noticeably, seven of the FA proteins form a multiprotein complex [49]. More broadly, interaction networks help identify genes with similar phenotypes [7,50]. For example, Lim et al. recently report an interaction network for human inherited ataxias that interconnect the majority of the ataxia-causing proteins and some modifier of neurodegeneration in animal models [51]. The ataxia interactome reveals common cellular pathways that might cause Purkinje cell dysfunction. It also provides an interesting tool for identifying new genes for inherited ataxias. The use of interaction networks to prioritize positional candidate disease genes identified by linkage or association studies is very promising approach that becomes more and more popular [52–54]. The most recent developments concern the integration of human phenotypes with protein interaction data to built a human phenome-interactome network [55]. More than 500 protein complexes associated with human genetic syndromes were deduced that were used to predict novel gene candidates for about 800 intervals linked to diseases.

5. Conclusion

After about a decade of intensive screening for protein interaction in various organisms, the efforts start to pay back. We witness a blooming number of publications reporting on the application of interactome network to human biology [56,57], structural biology [29] and human genetics [55]. These advances are linked to the recent availability of high quality interactomes maps. Despite this spectacular (encouraging) progress the road is still long until we achieve a comprehensive view on interaction networks. The available maps are still scanty and still suffer from a high rate of false negatives. Also, typically, they remain static and lack essential information capturing the logic of the molecular networks. They do not "qualify" the types of relationships between the various components. Because of their functional relevance, it is very likely that in the future interactions themselves (and not only the proteins involved) will deserve functional or spatial annotation.

Acknowledgements: Our work is supported by the EMBL and the EU-grant '3D repertoire'. We are grateful to John Briggs and Oriol Gallego for critical reading of the manuscript. We thank Toby Gibson for inspiring discussions.

References

[1] Lartigue, C., Glass, J.I., Alperovich, N., Pieper, R., Parmar, P.P., Hutchison 3rd., C.A., Smith, H.O. and Venter, J.C. (2007) Genome transplantation in bacteria: changing one species to another. Science 317, 632–638.
[2] Serrano, L. (2007) Synthetic biology: promises and challenges. Mol. Syst. Biol. 3, 158.
[3] Uetz, P. et al. (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. Nature 403, 623–627.
[4] Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M. and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc. Natl. Acad. Sci. USA 98, 4569–4574.
[5] Gavin, A.C. et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415, 141–147.
[6] Ho, Y. et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 415, 180–183.
[7] Gavin, A.C. et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440, 631–636.
[8] Krogan, N.J. et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 440, 637–643.
[9] Gingras, A.C., Gstaiger, M., Raught, B. and Aebersold, R. (2007) Analysis of protein complexes using mass spectrometry. Nat. Rev. Mol. Cell Biol. 8, 645–654.
[10] Shoemaker, B.A. and Panchenko, A.R. (2007) Deciphering protein–protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS Comput. Biol. 3, e43.
[11] Shoemaker, B.A. and Panchenko, A.R. (2007) Deciphering protein–protein interactions. Part I. Experimental techniques and databases. PLoS Comput. Biol. 3, e42.
[12] Vidal, M. (2005) Interactome modeling. FEBS Lett. 579, 1834–1838.
[13] Jeong, H., Mason, S.P., Barabasi, A.L. and Oltvai, Z.N. (2001) Lethality and centrality in protein networks. Nature 411, 41–42.
[14] Orchard, S. et al. (2007) The minimum information required for reporting a molecular interaction experiment (MIMIx). Nat. Biotechnol. 25, 894–898.
[15] Chen, T., Tsujimoto, N. and Li, E. (2004) The PWWP domain of Dnmt3a and Dnmt3b is required for directing DNA methylation to the major satellite repeats at pericentric heterochromatin. Mol. Cell. Biol. 24, 9048–9058.
[16] Anderson, J., Burns, H.D., Enriquez-Harris, P., Wilkie, A.O. and Heath, J.K. (1998) Apert syndrome mutations in fibroblast growth factor receptor 2 exhibit increased affinity for FGF ligand. Hum. Mol. Genet 7, 1475–1483.
[17] Pawson, T., Raina, M. and Nash, P. (2002) Interaction domains: from simple binding events to complex cellular behavior. FEBS Lett. 513, 2–10.
[18] Scheffzek, K. et al. (1997) The Ras-RasGAP complex: structural basis of GTPase activation and its loss in oncogenic Ras mutants. Science 277, 333–338.
[19] Schreiber, G. (2002) Kinetic studies of protein–protein interactions. Curr. Opin. Struct. Biol. 12, 41–47.
[20] Reichmann, D., Rahat, O., Cohen, M., Neuvirth, H. and Schreiber, G. (2007) The molecular architecture of protein–protein binding sites. Curr. Opin. Struct. Biol. 17, 67–76.
[21] Ladbury, J.E., Lemmon, M.A., Zhou, M., Green, J., Botfield, M.C. and Schlessinger, J. (1995) Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. Proc. Natl. Acad. Sci. USA 92, 3199–3203.
[22] Neduva, V. et al. (2005) Systematic discovery of new recognition peptides mediating protein interaction networks. PLoS Biol. 3, e405.
[23] Koch, C.A., Anderson, D., Moran, M.F., Ellis, C. and Pawson, T. (1991) SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. Science 252, 668–674.
[24] Meiyappan, M., Birrane, G. and Ladias, J.A. (2007) Structural basis for polyproline recognition by the FE65 WW domain. J. Mol. Biol. 372, 970–980.
[25] Neduva, V. and Russell, R.B. (2006) DILIMOT: discovery of linear motifs in proteins. Nucleic Acids Res. 34, W350–W355.

[26] Arkin, M. (2005) Protein–protein interactions and cancer: small molecules going in for the kill. Curr. Opin. Chem. Biol. 9, 317–324.
[27] Wells, J.A. and McClendon, C.L. (2007) Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. Nature 450, 1001–1009.
[28] Aloy, P., Ciccarelli, F.D., Leutwein, C., Gavin, A.C., Superti-Furga, G., Bork, P., Bottcher, B. and Russell, R.B. (2002) A complex prediction: three-dimensional model of the yeast exosome. EMBO Rep. 3, 628–635.
[29] Aloy, P. et al. (2004) Structure-based assembly of protein complexes in yeast. Science 303, 2026–2029.
[30] Alber, F. et al. (2007) Determining the architectures of macromolecular assemblies. Nature 450, 683–694.
[31] Alber, F. et al. (2007) The molecular architecture of the nuclear pore complex. Nature 450, 695–701.
[32] Heal, J.R., Roberts, G.W., Raynes, J.G., Bhakoo, A. and Miller, A.D. (2002) Specific interactions between sense and complementary peptides: the basis for the proteomic code. Chembiochem 3, 136–151.
[33] Gavin, A.C. and Superti-Furga, G. (2003) Protein complexes and proteome organization from yeast to man. Curr. Opin. Chem. Biol. 7, 21–27.
[34] Uetz, P. et al. (2006) Herpesviral protein networks and their interaction with the human proteome. Science 311, 239–242.
[35] Calderwood, M.A. et al. (2007) Epstein-Barr virus and virus human protein interaction maps. Proc. Natl. Acad. Sci. USA 104, 7606–7611.
[36] Charbonnier, S., Gallego, O. and Gavin, A.C. (2008) The social network of a cell: recent advances in interactome mapping. Biotechnol. Annu. Rev. 14, accepted for publication.
[37] Han, J.D. et al. (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. Nature 430, 88–93.
[38] Luscombe, N.M., Babu, M.M., Yu, H., Snyder, M., Teichmann, S.A. and Gerstein, M. (2004) Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431, 308–312.
[39] de Lichtenberg, U., Jensen, L.J., Brunak, S. and Bork, P. (2005) Dynamic complex formation during the yeast cell cycle. Science 307, 724–727.
[40] Xue, H. et al. (2007) A modular network model of aging. Mol. Syst. Biol. 3, 147.
[41] Li, C.Y., Mao, X. and Wei, L. (2008) Genes and (common) pathways underlying drug addiction. PLoS Comput. Biol. 4, e2.
[42] Hodgkin, J. (1998) Seven types of pleiotropy. Int. J. Dev. Biol. 42, 501–505.
[43] Dudley, A.M., Janse, D.M., Tanay, A., Shamir, R. and Church, G.M. (2005) A global view of pleiotropy and phenotypically derived gene function in yeast. Mol. Syst. Biol. 1, E1–E11.
[44] Egly, J.M. (2001) The 14th Datta Lecture. TFIIH: from transcription to clinic. FEBS Lett. 498, 124–128.
[45] Lehmann, A.R. (2001) The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. Genes Dev. 15, 15–23.
[46] Promislow, D.E. (2004) Protein networks, pleiotropy and the evolution of senescence. Proc. Biol. Sci. 271, 1225–1234.
[47] Krause, R., von Mering, C., Bork, P. and Dandekar, T. (2004) Shared components of protein complexes-versatile building blocks or biochemical artefacts? Bioessays 26, 1333–1343.
[48] Brunner, H.G. and van Driel, M.A. (2004) From syndrome families to functional genomics. Nat. Rev. Genet. 5, 545–551.
[49] Mace, G., Bogliolo, M., Guervilly, J.H., Dugas du Villard, J.A. and Rosselli, F. (2005) 3R coordination by Fanconi anemia proteins. Biochimie 87, 647–658.
[50] Walhout, A.J. et al. (2002) Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. Curr. Biol. 12, 1952–1958.
[51] Lim, J. et al. (2006) A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell 125, 801–814.
[52] Aerts, S. et al. (2006) Gene prioritization through genomic data fusion. Nat. Biotechnol. 24, 537–544.
[53] Franke, L., Bakel, H., Fokkens, L., de Jong, E.D., Egmont-Petersen, M. and Wijmenga, C. (2006) Reconstruction of a functional human gene network, with an application for prioritizing positional candidate genes. Am. J. Hum. Genet 78, 1011–1025.
[54] Oti, M., Snel, B., Huynen, M.A. and Brunner, H.G. (2006) Predicting disease genes using protein–protein interactions. J. Med. Genet. 43, 691–698.
[55] Lage, K. et al. (2007) A human phenome-interactome network of protein complexes implicated in genetic disorders. Nat. Biotechnol. 25, 309–316.
[56] Chuang, H.Y., Lee, E., Liu, Y.T., Lee, D. and Ideker, T. (2007) Network-based classification of breast cancer metastasis. Mol. Syst. Biol. 3, 140.
[57] Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M. and Barabasi, A.L. (2007) The human disease network. Proc. Natl. Acad. Sci. USA 104, 8685–8690.
[58] Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. J. Mol. Graph. 14, 33–38, 27–28.